

# Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics

Taro Takaguchi<sup>1</sup>, Naoki Masuda<sup>1</sup>, and Petter Holme<sup>2,3,4,\*</sup>

<sup>1</sup> Department of Mathematical Informatics,  
The University of Tokyo,  
7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan

<sup>2</sup> IceLab, Department of Physics, Umeå University,  
901 87 Umeå, Sweden

<sup>3</sup> Department of Energy Science, Sungkyunkwan University,  
Suwon 440-746, Korea

<sup>4</sup> Department of Sociology, Stockholm University,  
106 91 Stockholm, Sweden

\* Corresponding author (petter.holme@physics.umu.se)

March 7, 2013

## Abstract

Records of social interactions provide us with new sources of data for understanding how interaction patterns affect collective dynamics. Such human activity patterns are often bursty, *i.e.*, they consist of short periods of intense activity followed by long periods of silence. This burstiness has been shown to affect spreading phenomena; it accelerates epidemic spreading in some cases and slows it down in other cases. We investigate a model of history-dependent contagion. In our model, repeated interactions between susceptible and infected individuals in a short period of time is needed for a susceptible individual to contract infection. We carry out numerical simulations on real temporal network data to find that bursty activity patterns facilitate epidemic spreading in our model.

# 1 Introduction

Communication between individuals is a fundament of human society. Nowadays technologies such as sensor devices and online communication services provide us with records of interaction between individuals, including face-to-face conversations, e-mail exchanges, and phone calls, in massive amounts. Such data often consist of a sequence of interaction events. Each event is represented by a triplet, *i.e.*, the IDs of two individuals involved in the event and the time of the event. One traditional way to characterize such data is to represent them as an aggregated network, in which the links are drawn between two nodes (*i.e.*, individuals) that communicate in at least one event, and investigate structural properties of the aggregated static networks [1]. Another and richer representation of this type of data is to model them as temporal networks, in which the links between two nodes exist only at the time of an event [2].

Effects of temporal networks on contagious phenomena, such as infectious diseases and rumors, have been investigated by various authors. To simulate spreading dynamics on temporal networks, we read the events in an empirical event sequence one by one in the chronological order and possibly update the states (*e.g.*, susceptible and infected) of the two nodes involved in the event. Karsai and colleagues simulated the susceptible-infected (SI) model on temporal networks and found that bursty activity patterns slow down contagions [3]; Bursty activity patterns are identified with a long-tailed distribution of the interevent intervals (IEIs) [4, 5]. The slowing down occurs because, at an arbitrary time point, the average time to the next event is longer for the long-tailed IEI distribution than for the exponential IEI distribution with the same mean. In other words, after an individual gets infected, it tends to take longer time to infect the neighbors under the long-tailed as compared to exponential IEI distribution. Other numerical [6, 7] and analytical [8–10] results also support that the long-tailed IEI distribution mitigates contagion. However, the burstiness was reported to accelerate contagion on a different data set [11] and a different type of epidemic dynamics [12]. Our understanding of the effect of the burstiness on contagious processes is still elusive.

In the present study, we show that bursty activity patterns facilitate epidemic spreading in a variant of the deterministic threshold model [13, 14]. In standard models of epidemics including the SI, susceptible-infected-recovered (SIR), and susceptible-exposed-infected-recovered (SEIR) models, which have been employed in the literature cited above, a susceptible node gets infected from an infected neighbor with a constant probability in an event, regardless of the amount of exposure to infected neighbors in the past. However, history-dependent thresholding effects in which the thresholding operates on the concentration of the pathogen have been reported for some infectious diseases mediated by bacteria, such as the tuberculosis and the dysentery [15]. In the case of information propagation, the exposure to the information increases one’s interest in a topic, and the attractiveness of a topic decays in time in the absence of stimulus [16, 17]. We may need multiple interactions to persuade others to do something, and repeated contacts in a short period can be more effective than those dispersed over a long period. To consider this type of infection, we generalize the deterministic threshold model to the case of history dependence and memory decay and simulate the proposed model on temporal network data.

## 2 Methods

Each node  $i$  is assumed to have an internal variable denoted by  $v_i \geq 0$  ( $i = 1, 2, \dots, N$ ), which represents, for example, the concentration of a pathogen in the individual or the individual’s interest in a topic. Initially,  $v_i$  to equal to zero for all  $i$ . We assume that node  $i$  is in the susceptible ( $S$ ) state before  $v_i$  exceeds a threshold value  $v_{\text{thr}}$  and that node  $i$  is in the infected ( $I$ ) state once  $v_i$  exceeds  $v_{\text{thr}}$ . Each node is in either state. Nodes in state  $I$  never return to state  $S$ ; our model is an extension of the SI model. Therefore, the number of  $I$  nodes monotonically increases in time.

When node  $i$  in state  $S$  interacts with an  $I$  node through an event,  $v_i$  is increased by unity. In the absence of interaction with  $I$  nodes,  $v_i$  is assumed to decay exponentially in time. In

other words,  $v_i$  is given by

$$v_i(t) = \sum_{t_e} \exp\left(-\frac{t - t_e}{\tau_d}\right), \quad (1)$$

where  $t_e$  is the time of an event between node  $i$  and an  $I$  node, and  $\tau_d$  is the decay time constant. An example time course of  $v_i(t)$  is shown in Fig. 1.

The model contains two parameters  $\tau_d$  and  $v_{\text{thr}}$  and can be regarded as a variant of the deterministic threshold model [13, 14]. Although we assume that all the nodes have the same values of  $\tau_d$  and  $v_{\text{thr}}$  for simplicity, it is straightforward to generalize the model in the case of heterogeneous parameter values.

We simulate our model numerically on empirical temporal networks in the following way. At  $t = 0$ , we select a node as initial seed and set its state to  $I$ . All the other nodes are initially in state  $S$ . Then, we chronologically read the event sequence one by one and update  $v_i$  and the states of the two nodes involved in the event. Because our model is deterministic, the final infection size (*i.e.*, fraction of  $I$  nodes at time  $t_{\text{max}}$ , where  $t_{\text{max}}$  is the time of the last event in the data set), denoted by  $I_i$ , is unique for given initial seed  $i$ ,  $\tau_d$ , and  $v_{\text{thr}}$ .

We use two data sets. The first data set, called Conference in the following, is the face-to-face conversation log between attendees of a scientific conference [18]. The second data set, called Email, is the record of e-mail exchanges between the members of a university [19]. In the second data set, we neglect the direction of the interaction (*i.e.*, from sender to receiver) for simplicity. The basic statistics of the data sets are summarized in Tab. 1.

### 3 Results

In Figs. 2(a) and 2(b), we plot the dependence of final infection size  $I_m$  on  $\tau_d$  and  $v_{\text{thr}}$  for initial seed node  $m$  having the maximum number of events in Conference and Email data sets, respectively. In the blank parameter region, no infection occurs such that  $I_m = 1/N$ . Naturally,  $I_m$  increases with  $\tau_d$  and decreases with  $v_{\text{thr}}$ .

Next, we carry out the same set of simulations on the randomized temporal networks for the

sake of comparison. To this end, we use the so-called randomly-permuted-times randomization, in which the time stamps of all the events are randomly shuffled [2, 3, 6]. The randomization eliminates temporal properties of the original temporal networks such as bursty activity patterns, daily and weekly patterns, and the pairwise correlations of the IEs, whereas it conserves all the properties of the aggregated networks, *i.e.*, weighted adjacency matrix.

For the randomized temporal networks, the dependence of  $I_m$  on  $\tau_d$  and  $v_{\text{thr}}$  are shown in Figs. 2(c) and 2(d) for Conference and Email data sets, respectively. We find that the parameter region in which infection occurs is larger for the original temporal networks (colored regions in Figs. 2(a) and 2(b)) than for the randomized temporal networks (colored regions in Figs. 2(c) and 2(d)) for intermediate values of  $\tau_d$  ( $10^2 \leq \tau_d \leq 10^4$  and  $10^4 \leq \tau_d \leq 10^6$  for Conference and Email data sets, respectively). In the original data sets, the nodes tend to have many events in bursty periods and be quiescent in other periods. The randomization procedure eliminates bursty activity patterns. Therefore,  $v_m(t)$  can reach  $v_{\text{thr}}$  in such a bursty period for the original but not randomized temporal networks if  $\tau_d$  and  $v_{\text{thr}}$  take intermediate values. In the randomized data sets,  $v_m(t)$  tends to decay faster than it grows, although the number of events per node is the same between the original and randomized data.

For Email data set,  $I_m$  for the randomized data set (Fig. 2(d)) is larger than that for the original data set (Fig. 2(c)) when  $\tau_d$  is large and  $v_{\text{thr}}$  is small. This is mainly because the randomization increases the reachability ratio from initial seed  $m$  to a large extent. The reachability ratio from a node is defined as the fraction of nodes that we can reach from the node by tracing the events in the chronological order [20]. If every event can elicit infection, which is the case when  $\tau_d$  is large and  $v_{\text{thr}}$  is small,  $I_m$  is approximated by the reachability ratio from node  $m$ . The reachability ratio from node  $m = 3024$  in Email data set is equal to 0.7458 and 0.9981 for the original and randomized data sets, respectively. In contrast, the reachability ratio from node  $m = 55$  in Conference data set is equal to 0.9642 and 1 for the original and randomized data sets, respectively; the difference is smaller than in the case of Email data set.

In Fig. 3, the average final infection size  $\langle I_i \rangle$ , defined as the average of  $I_i$  over all the nodes  $i$ ,

is plotted as a function of  $\tau_d$  for two values of  $v_{\text{thr}}$  for each data set. Figure 3 indicates that  $\langle I_i \rangle$  for the original temporal networks is larger than that for the randomized temporal networks for a broad range of  $\tau_d$  for both data sets.

In the bond percolation on static networks, the probability that single bonds are open (independent of different bonds) is the sole parameter that determines the possibility that the entire network has a giant component [1]. Motivated by this picture, we hypothesize that the results shown in Figs. 2 and 3 are largely explained by the bursty nature of events on single links. In other words, we speculate that the structure of the aggregated networks or correlation between event sequences on different links do not much influence the results. To test the hypothesis, we separately examine the event sequence on each link. For each link, *i.e.*, node pair  $(i, j)$  with at least one event,  $T_{j \rightarrow i}(t)$  is defined as the time required for node  $i$  to be infected since node  $j$  has been infected. We emphasize that we do not consider influences from other nodes on  $i$  in this analysis. We take the time average of  $T_{j \rightarrow i}(t)$ , denoted by  $\overline{T}_{j \rightarrow i}$ , over  $0 \leq t \leq t_{\text{max}}$ . A problem with the time averaging is that  $T_{j \rightarrow i}(t)$  is indefinite for sufficiently large  $t$  because  $i$  does not get infected by time  $t_{\text{max}}$ . Therefore, we adopt the boundary condition in which the first events between nodes  $i$  and  $j$  virtually replay after  $t = t_{\text{max}}$ . We denote the time of the first event between  $i$  and  $j$  by  $t_1$ . If we temporarily set  $T_{j \rightarrow i}(t_{\text{max}} + t_1) = T_{j \rightarrow i}(t_1)$ , it takes at most  $t_{\text{max}} - t + t_1 + T_{j \rightarrow i}(t_1)$  for node  $i$  starting with  $v_i(t) = 0$  to be infected from node  $j$ , where  $t_{\text{last}} \leq t \leq t_{\text{max}}$  and  $t_{\text{last}}$  is the last time before which  $T_{j \rightarrow i}(t)$  is finite. Therefore, we set  $T_{j \rightarrow i}(t) = t_{\text{max}} - t + t_1 + T_{j \rightarrow i}(t_1)$  for  $t_{\text{last}} \leq t \leq t_{\text{max}}$ . This boundary condition is the same as that is used in Ref. [21] for defining the average temporal path length. If  $T_{j \rightarrow i}(t_1)$  is indefinite (*i.e.*, infection never occurs between  $i$  and  $j$ ),  $\overline{T}_{j \rightarrow i}$  is set to infinite. We define denoted by  $\langle 1/\overline{T}_{j \rightarrow i} \rangle$  as the average of  $1/\overline{T}_{j \rightarrow i}$  over the 20% links with the largest numbers of events, because the majority of the links possesses a small number of events in both data sets. This thresholding leaves 441 and 6,932 links for Conference and Email data sets, respectively.

$\langle 1/\overline{T}_{j \rightarrow i} \rangle$  for the original and randomized temporal networks are shown for various  $\tau_d$  and  $v_{\text{thr}}$  values for Conference (Figs. 4(a) and 4(b)) and Email (Figs. 4(c) and 4(d)) data sets.

Because infection can be induced only through a single link in the present simulations, we examined  $v_{\text{thr}}$  values that are much smaller than those used in Figs. 2 and 3. For both data sets,  $\langle 1/\overline{T}_{j \rightarrow i} \rangle$  for the original temporal networks (Figs. 4(a) and 4(c)) is larger than that for the randomized networks (Figs. 4(b) and 4(d)) for intermediate values of  $\tau_d$  ( $10^2 \leq \tau_d \leq 10^4$  and  $10^4 \leq \tau_d \leq 10^6$  for Conference and Email data sets, respectively). The behavior of  $\langle 1/\overline{T}_{j \rightarrow i} \rangle$  is consistent with the results of the network-based simulations (Figs. 2 and 3).

## 4 Conclusions

We numerically simulated a variant of the deterministic threshold model on empirical temporal networks. We found that the average final infection size for the empirical temporal networks is larger than those for the randomized temporal networks in a broad parameter region (Figs. 2 and 3). The bursty nature of the IEs on single links has a sufficient explanatory power for the results of the network-based simulations (Fig. 4). The burstiness promoted epidemic spreading when the decay exponent  $\tau_d$  takes an intermediate value ( $10^2 \leq \tau_d \leq 10^4$  and  $10^4 \leq \tau_d \leq 10^6$  (seconds) for Conference and Email data sets, respectively). This range of  $\tau_d$  may be practical because the influence of a pathogen that an individual has received may last for hours to days.

The finding that the burstiness facilitates the spreading also sheds light on a function of the redundant interaction events. We previously found that about 80% of the events are redundant in the sense that they affect little on bridging efficient temporal paths in Conference data set [22]. However, for the spreading dynamics in our model, such redundant events play a crucial role in increasing  $v_i(t)$  within bursty periods.

## Acknowledgments

The authors thank to the SocioPatterns collaboration (<http://www.sociopatterns.org>) for providing the data set. This research is supported by the Aihara Project, the FIRST program from JSPS, initiated by CSTP. T.T. acknowledges support provided through Grants-in-Aid for



Scientific Research (No. 10J06281) from JSPS, Japan. N.M. acknowledges support provided through Grants-in-Aid for Scientific Research (No. 23681033) from MEXT, Japan. P.H. acknowledges support from the Swedish Research Council and the WCU program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology R31-2008-10029.

## References

- [1] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
- [2] P. Holme and J. Saramäki, preprint arXiv:1108.1780.
- [3] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, Phys. Rev. E **83**, 025102(R) (2011).
- [4] A.-L. Barabási, Nature **435**, 207–211 (2005).
- [5] A. Vázquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási, Phys. Rev. E **73**, 036127 (2006).
- [6] G. Miritello, E. Moro, and R. Lara, Phys. Rev. E **83**, 045102(R) (2011).
- [7] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Regis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, and P. Vanhemss, BMC Med. **9**, 87 (2011).
- [8] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási, Phys. Rev. Lett. **98**, 158702 (2007).
- [9] J. L. Iribarren and E. Moro, Phys. Rev. Lett. **103**, 038702 (2009).
- [10] B. Karrer and M. E. J. Newman, Phys. Rev. E **82**, 016101 (2010).
- [11] L. E. C. Rocha, F. Liljeros, and P. Holme, PLoS Comput. Biol. **7**, e1001109 (2011).
- [12] F. Karimi and P. Holme, unpublished (2012).

- [13] P. S. Dodds and D. J. Watts, Phys. Rev. Lett. **92**, 218701 (2004).
- [14] P. S. Dodds and D. J. Watts, J. Theor. Biol. **232**, 587–604 (2004).
- [15] R. I. Joh, H. Wang, H. Weiss, and J. S. Weitz, Bull. Math. Biol. **71** 845–862 (2009).
- [16] R. Crane and D. Sornette, Proc. Natl. Acad. Sci. USA **105**, 15649–15653 (2008).
- [17] L. Averell and A. Heathcote, J. Math. Psychol. **55**, 25–35 (2011).
- [18] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, J. Theor. Biol. **271**, 166–180 (2011).
- [19] J.-P. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. USA **101**, 14333–14337 (2004).
- [20] P. Holme, Phys. Rev. E **71**, 046119 (2005).
- [21] R. K. Pan and J. Saramäki, Phys. Rev. E **84**, 016105 (2011).
- [22] T. Takaguchi, N. Sato, K. Yano, and N. Masuda, preprint arXiv:1205.4808.

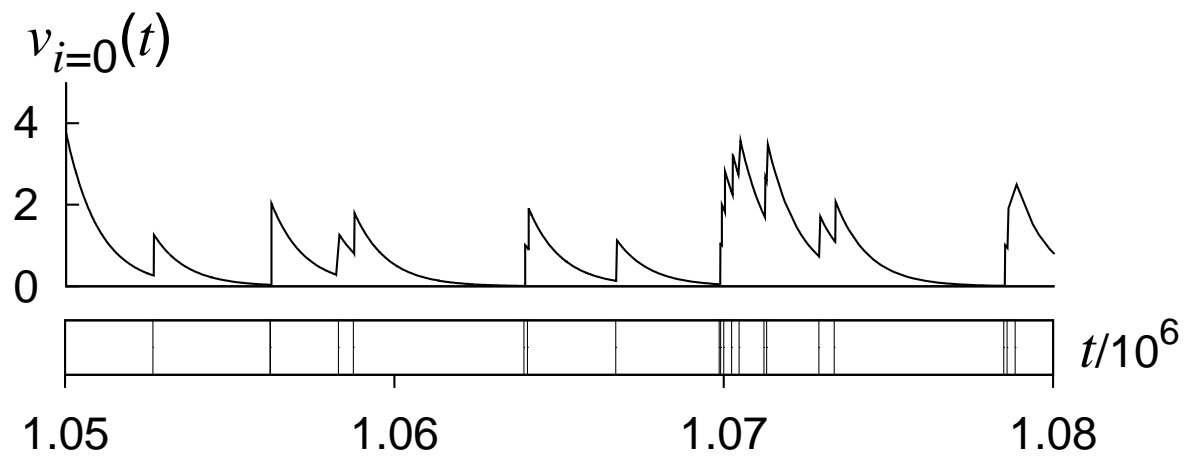


Figure 1:  $v_{i=0}(t)$  for  $1.05 \times 10^6 \leq t \leq 1.08 \times 10^6$  in Email data set. We set  $\tau_d = 1000$ . The vertical ticks in the box plot in the bottom indicate the times of the events that involve node  $i = 0$ .

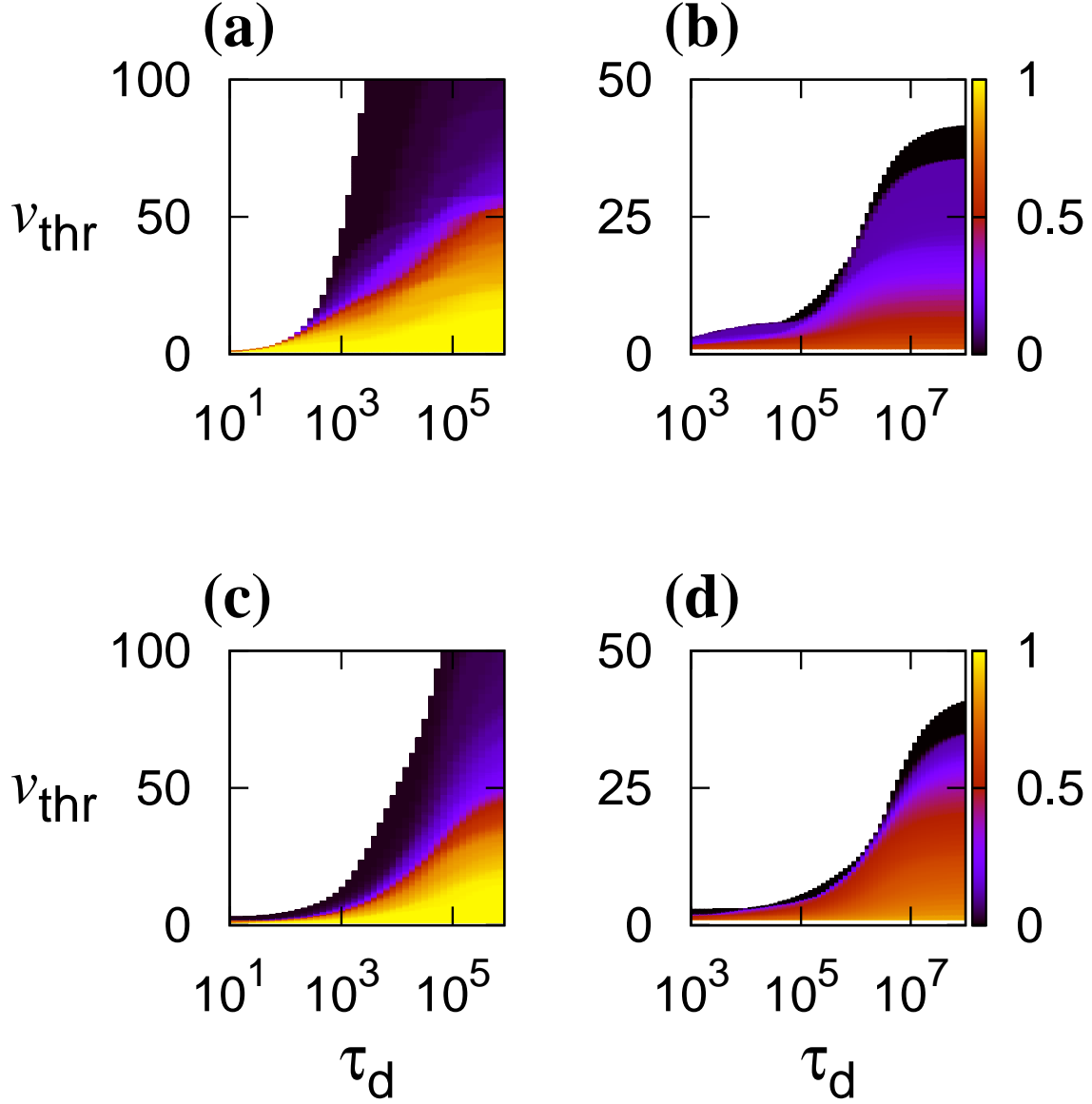


Figure 2: Dependence of the final infection size  $I_m$  on  $\tau_d$  and  $v_{\text{thr}}$ . (a), (b) Original temporal networks. (c), (d) Randomized temporal networks. (a), (c)  $I_{m=55}$  in Conference data set. (b), (d)  $I_{m=3024}$  in Email data set. No infection occurs in the black parameter regions. The parameter values for which at least one infection occurs are colored.

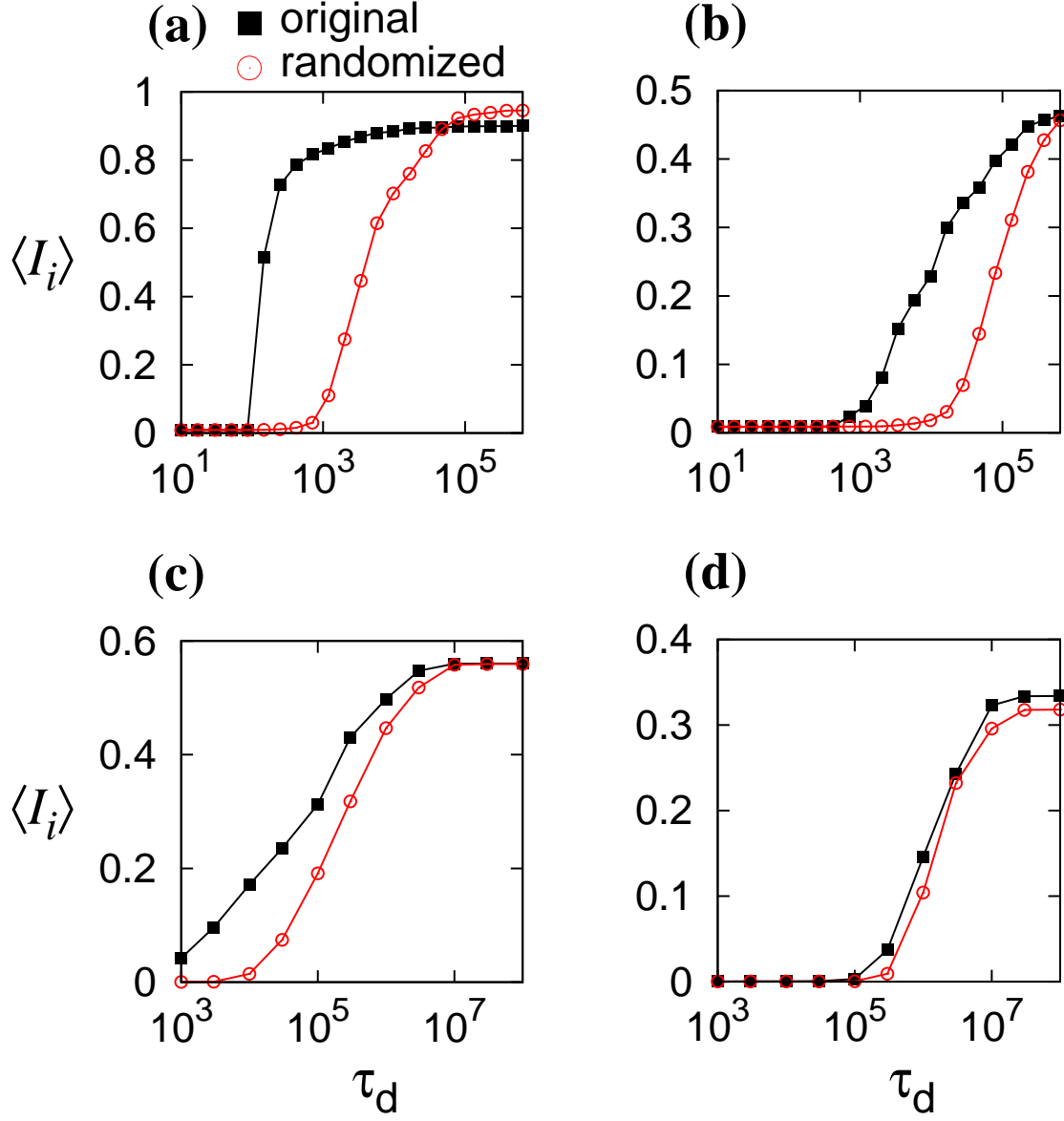


Figure 3: Average final infection size  $\langle I_i \rangle$  for (a, b) Conference and (c, d) Email data sets. Squares and circles correspond to the original and randomized temporal networks, respectively. We set (a)  $v_{\text{thr}} = 5$ , (b)  $v_{\text{thr}} = 20$ , (c)  $v_{\text{thr}} = 3$ , and (d)  $v_{\text{thr}} = 10$ .

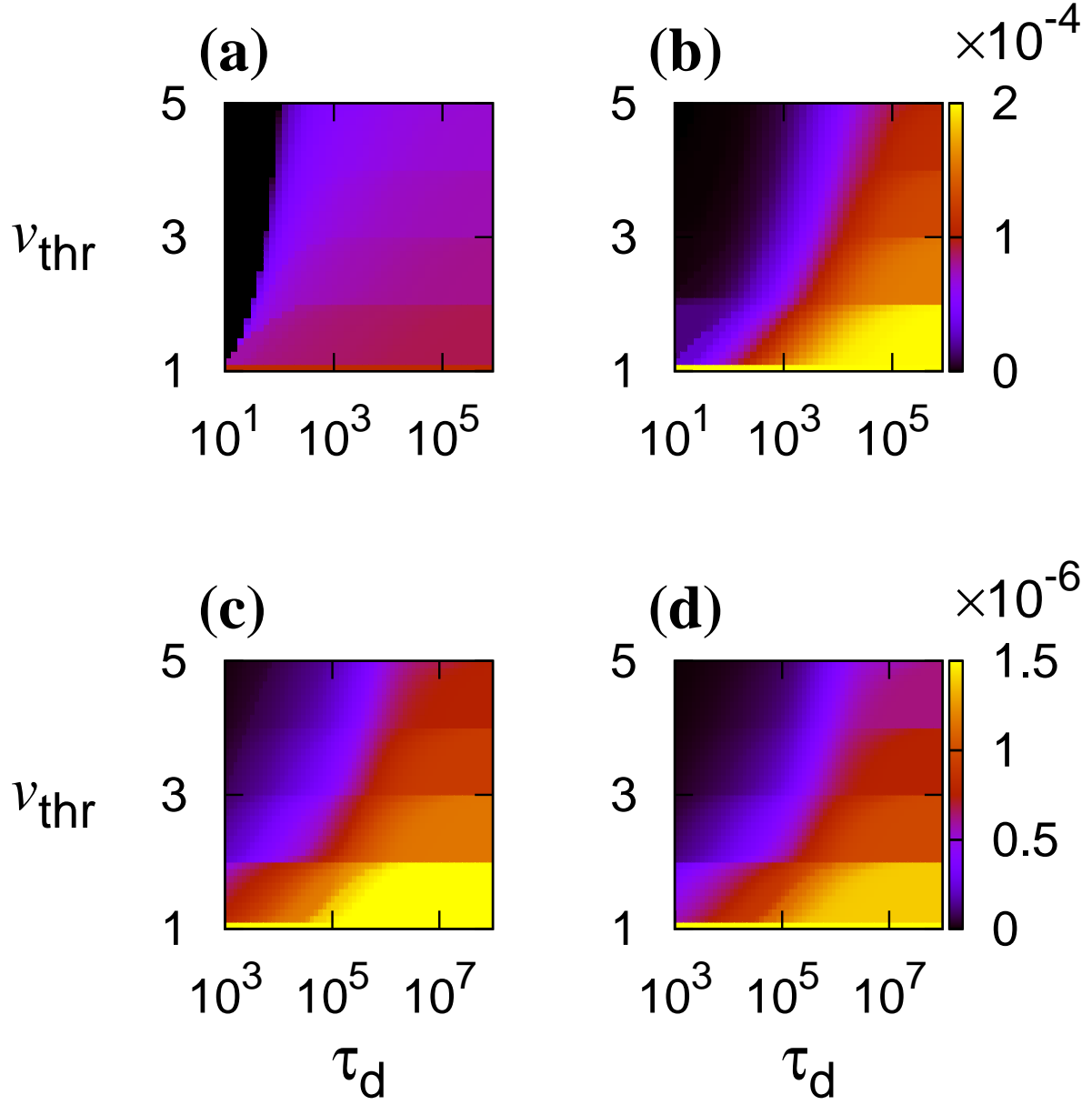


Figure 4: Average single-link infection rate  $\langle 1/\overline{T}_{j \rightarrow i} \rangle$  for (a), (b) Conference and (c), (d) Email data sets. (a), (c) Original temporal networks. (b), (d) Randomized temporal networks.

Table 1: Statistics of the two data sets.

	Conference	Email
Number of nodes ( $N$ )	113	3,188
Number of events	20,808	309,125
Recording period	3 days	83 days
Time resolution	20 sec	1 sec